日本語機能表現辞書の編纂

松吉　俊[†,††]・佐藤　理史[††]・宇津呂武仁[†††]

日本語には，「にたいして」や「なければならない」に代表されるような，複数の形態素からなっているが，全体として１つの機能語のように働く複合辞が多く存在する．われわれは，機能語と複合辞を合わせて機能表現と呼ぶ．本論文では，自然言語処理のための日本語機能表現辞書について提案する．日本語の機能表現が持つ主な特徴の１つは，個々の機能表現に対して，多くの異形が存在することである．計算機が利用することを想定した辞書を編纂する場合，これらの異形を適切に扱う必要がある．われわれが提案する辞書は，機能表現の異形を体系的に整理するために，見出し体系として，９つの階層からなる階層構造を用いる．現在，この辞書には，341の見出し語と16,771の出現形が収録されており，既存の機能表現リストと比較した結果，各々の見出し語に対して，ほぼすべての異形が網羅されていることが確かめられた．

キーワード：辞書，複合語表現，機能語，複合辞

A Dictionary of Japanese Functional Expressions with Hierarchical Organization

Suguru Matsuyoshi[†,††], Satoshi Sato[††] and Takehito Utsuro[†††]

The Japanese language has a lot of functional expressions, each of which consists of more than one word and behaves like a single function word. A remarkable characteristic of Japanese functional expressions is that each functional expression has many different surface forms. This paper proposes a dictionary of Japanese functional expressions with hierarchical organization. We use a hierarchy with nine abstraction levels: the root node is a dummy node that governs all entries; a node in the first level is a headword in the dictionary; a leaf node corresponds to a surface form of a functional expression. We have compiled the dictionary with 341 headwords and 16,771 surface forms, which covers almost all of surface forms for each headword.

Key Words: *dictionary, multi-word expressions, function words, complex particles*

† 京都大学大学院情報学研究科, Graduate School of Informatics, Kyoto University
†† 名古屋大学大学院工学研究科, Graduate School of Engineering, Nagoya University
††† 筑波大学大学院システム情報工学研究科, Graduate School of Systems and Information Engineering, University of Tsukuba

1 はじめに

日本語の解析システムは，1990年代にそれまでの研究が解析ツールとして結晶し，現在では，各種の応用システムにおいて，それらの解析ツールが入力文を解析する解析モジュールとして利用されるようになってきている．解析ツールを利用した応用システムの理想的な構成は，与えられた文を解析する解析ツールと，その後の処理を直列につなげた，図1に示すような構成である．例えば，情報抽出システムでは，応用モジュールは，解析ツールの出力データを受け取り，そのデータに抽出すべき情報が含まれているかどうかを調べ，含まれている場合にその情報を抽出する，という処理を行なうことになる．

ここで，応用モジュールを，おおきく，次の2つのタイプに分類する．

(1) **言語表現そのもの（言語構造）を対象とする**応用モジュール
例えば，目的格と述語の組を認識して，それらの数を数えるモジュール．

(2) **言語表現が伝える情報（情報構造）を対象とする**応用モジュール
例えば，「どのメーカーが，なんという製品を，いつ発売するか」という情報を抽出する情報抽出モジュール．

後者のモジュールでは，どのような言語表現が用いられているのかが問題となるのではなく，どのような言語表現が用いられていようと，それが「どのメーカーが，なんという製品を，いつ発売するか」という情報を伝達しているのであれば，それを抽出することが要求される．われわれが想定する応用モジュールは，この後者のタイプである．

応用システムにおいて，解析ツールは，「応用に特化しない言語解析処理をすべて担う」ことを期待される．しかしながら，現実は，そのような理想的な状況とは程遠く，応用モジュールを構築する際に，現在の解析ツールが放置しているいくつかの言語現象と向き合うことを余儀なくされる．そのような言語現象の具体例は，おおきく，以下の4種類に分類できる．

表記の問題
いわゆる「表記のゆれ」が放置されているので，これらを同語とみなす処理が必要となる．例えば，「あいまい」と「曖昧」，あるいは，「コンピューター」と「コンピュータ」．

単位の問題
複合語表現 (multi-word expressions) の設定が不十分であるので，追加認定を行なう処理が必要となる．

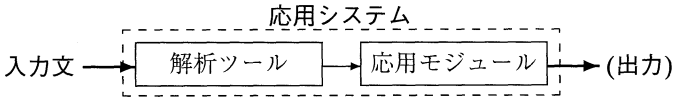

図 1 応用システムの理想的な構成

外部情報源とのインタフェースの問題

語の認定が行なわれないので，他の外部情報源を利用する場合，文字列でインタフェースをとるしか方法がない．それぞれの情報源（例えば，一般の国語辞典）で，品詞体系や見出し表記が異なるので，かなりの辞書参照誤りが発生する．

異形式同意味の問題

言語表現は異なるが伝達する情報（意味）が同じものが存在するので，これらを同一化する処理が必要となる．

これらの問題に共通するキーワードは，「情報（意味）の基本単位」である．

日本語の表現は，内容的・機能的という観点から，おおきく2つに分類できる．さらに，「表現を構成する語の数」という観点を加えると，表1のように分類できる．ここで，**複合辞**とは，「にたいして」や「なければならない」のように，複数の語から構成されているが，全体として1つの機能語のように働く表現のことである．われわれは，機能的というカテゴリーに属する機能語と複合辞を合わせて**機能表現**と呼ぶ．

内容的表現に関しては，近年，上の4つの問題を解決するための研究が行なわれている．例えば，内容語に関しての研究 (佐藤 2004; 黒橋，河原 2005; 浅原他 2005) や慣用表現に関しての研究 (尾嶋他 2006; 橋本他 2006a, 2006b) がある．その一方で，機能表現に関しては，大規模な数のエントリーに対して上記の問題を解決しようとする研究はほとんど存在しない．大規模な数の機能表現を扱ったものに，Shudo ら (Shudo et al. 2004) や兵藤ら (兵藤他 2000) による研究があるが，それらは，上記の問題を考慮していない．

このような背景により，本研究では，自然言語処理において日本語機能表現を処理する基礎となるような**日本語機能表現辞書**を提案する．この辞書は，大規模な数の機能表現に関して，上記の問題に対する1つの解決法を示す．

本論文は，以下のように構成される．まず，第2章で，機能表現とその異形について述べる．次に，第3章において，日本語機能表現辞書の設計について述べる．第4章で，辞書の見出し体系として採用した，機能表現の階層構造について説明する．そして，第5章で，辞書の編纂手順について説明し，現状を報告する．第6章で，関連研究について述べ，最後に，第7章でまとめを述べる．

表 1　日本語表現の分類

	一語	二語以上
内容的	内容語（名詞，動詞，形容詞など）	複合名詞，複合動詞，慣用表現
機能的	**機能語**（助詞，助動詞，接続詞，形式名詞）	**複合辞**

2 日本語機能表現

2.1 複合辞

複合辞というとらえ方を初めて提唱したのは，国立国語研究所の資料 (山崎，藤田 2001) によると，永野 (永野 1953) である．永野は，語源的・構造的にはいくつかの語に分解できるが，単なる部分の合成以上の「一まとまりの意味を持っているものと見てよい」連語表現の存在を指摘した．例えば，「からには」は，3つの助詞「から」，「に」，「は」から構成されているが，それは，全体として「特に理由を提示して，課題の場を設定し，次に来る陳述を強く期待させる場合に使われる言い方」であると説明する．彼は，このような表現を複合助詞と呼んだ．そして，同様の基準で複合助動詞，複合接続詞，複合感動詞についても考え，複合助詞とこれらを一まとめにして，**複合辞**と呼んだ．永野は，内容語の領域において，複合語やイディオムが表現資材としての単位であるのと同様に，複合辞を表現上の単位と考える．

2.2 機能表現の定義

機能表現とは，おおよそ，文中においてなんらかの機能を持つ表現ということができる．しかし，機能表現全体に対して，「なんらかの機能」をより精密に書き下すことはほとんど不可能であると考えられるので，本研究では，機能表現の定義として，次のような定義を採用する．

> 表2に示すいずれかの機能を持つ表現を，**それぞれの機能型に属する機能表現**と呼び，その総称として**機能表現**という用語を用いる．

表2に挙げた機能は，通常，機能語に対して認定されるものである．本研究では，それを複合辞にまで拡張し，機能表現の定義として用いることにした．

上記の定義を用いると，「からには」や「ばかりか」など，機能語の列からなる表現や，「なければならない」や「かもしれない」など，自立性が低い内容語を含む表現を機能表現であると判断することができ，「じゅんに」や「経過してから」など，内容語の自立性が高く，一まとまり

表 2 機能と L^3ID (4章参照)

機能	機能型	例	L^3ID
前件を後件の用言に関係付ける	格助詞型	について	P
前件を後件の節に関係付ける	接続助詞型	にもかかわらず	Q
前件を後件の体言に関係付ける	連体助詞型	にたいする	D
前の文を後ろの文に関係付ける	接続詞型	ところが	C
前件に付加的なニュアンスを与える	助動詞型	なければならない	M
前件を名詞化する	形式名詞型	こと	N
前件を取り立てる	とりたて詞型	のみならず	T
前件を話題化する	提題助詞型	といえば	W

の表現とはみなしにくいものを機能表現ではないと判断することができる．しかしながら，一般に，ある表現が機能表現であるかどうかの判断は，それほど容易ではない．その理由は，次の2点に対して統一した見解が存在しないことにある．

機能表現と呼べる表現の範囲

「にもとづいて」や「と比較して」など，機能表現であるかどうかの判断が難しい表現が存在する．これらの表現が機能表現であるかどうかについては，上記の定義や，これまでに機能表現であると判断された表現などを参照し，主観的に判断することになる．

機能表現の単位

「たにちがいない」，「たとしても」，「ないかもしれなかった」など，全体で1単位の機能表現とみなすべきか，それとも，複数の機能表現からなる表現であるとみなすべきか判断が揺れる表現が存在する．

本研究では，前者に対して保守的な立場をとり，定評のある文献において機能表現であると認められているもののみを機能表現と認め，機能表現であるかどうかの判断が難しい表現に対しては，機能表現と認めることを保留する．一方，後者に対しては，不必要に長い機能表現を認めるのではなく，例えば，「ないかもしれなかった」に対しては，「ない」，「かもしれなかっ」，「た」のように，適切な構成要素に分割し，その各々を機能表現であるとみなす．

2.3 種々の異形

日本語の機能表現が持つ主な特徴の1つは，個々の機能表現に対して，多くの**異形**が存在することである．例えば，「なければならない」に対して，「なくてはならない」，「なくてはならず」，「なければなりません」，「なけりゃならない」，「なければならぬ」，「ねばならん」など，多くの異形が存在する．このような異形をつくり出す過程には，いくつかの言語現象が絡んでいる．われわれは，これらの言語現象に基づいて，機能表現の異形を次の7つのカテゴリーに分類した．

(1) **派生**

2つの表現がお互いに緊密に関連しているが，それらの機能が異なるとき，われわれは，それらを派生に分類する．例えば，「にたいして」と「にたいする」は，いずれも格助詞「に」と動詞「たいする」の1つの活用形という形態をしており，お互いに緊密に関連している．その一方で，それらは，異なる機能を持っている．「にたいして」は，格助詞型機能表現であり，「にたいする」は，連体助詞型機能表現である．それゆえに，これらを派生に分類する．

(2) **機能語の交替**

機能表現を構成する機能語が，異なる機能語に置き換えられることにより，異形が生成されることがある．例えば，「からすると」の末尾の「と」を「ば」に置き換えると，「か

らすれば」という異形が生成される．

(3) **音韻的変化**

機能表現の構成要素が音韻的に変化することにより，異形が生成されることがある．音韻的変化は，次の 4 種類に分類できる．

(a) **縮約**

特定の文字列が縮約することにより，異形が生成されることがある．例えば，「なければならない」の「れば」が「りゃ」へ縮約した場合，「なけりゃならない」という異形が生成される．

(b) **脱落**

特定の文字が脱落することにより，異形が生成されることがある．例えば，「ところだった」から「ろ」が脱落することにより，「とこだった」という異形が生成される．

(c) **促音化・撥音化**

特定の文字列が促音化，もしくは撥音化することにより，異形が生成されることがある．例えば，「たものではない」の「の」が撥音化することにより，「たもんではない」という異形が生成される．

(d) **有声音化**

前に接続する語により，機能表現の先頭の子音が有声音になり，異形が生成されることがある．例えば，「ていい」は，前に「読む」が接続する場合，先頭の子音 “t” が有声音 “d” になり，「でいい」という異形が生成される．

(4) **とりたて詞の挿入**

機能表現の内部にとりたて詞 (沼田 1986) が挿入されることにより，異形が生成されることがある．例えば，「といっても」の「と」と「いっても」の間には，とりたて詞「は」が挿入可能である．この挿入により，「とはいっても」という異形が生成される．

(5) **活用**

機能表現を構成する末尾の語が活用することにより，異形が生成されることがある．例えば，「なければならない」の末尾の「ない」が「なかっ」に活用することにより，「なければならなかっ」という異形が生成される．

(6) **「です/ます」の有無**

機能表現の内部に「です」や「ます」が挿入されることにより，異形が生成されることがある．例えば，「にたいして」の「にたいし」と「て」の間には，「ます」が活用した「まし」が挿入可能である．この挿入により，「にたいしまして」という異形が生成される．

(7) **表記のゆれ**

機能表現の構成語が漢字表記を持っている場合，その語の表記の仕方によって，異形が

生成されることがある．例えば，「にあたって」に対して，「に当たって」，「に当って」という漢字表記の異形が存在する．

3　機能表現辞書の設計

3.1　辞書に求める要件

機能表現の辞書を作成するためには，次の2種類のリストが必要であると考えられる．

(1)　辞書の見出し語のリスト

(2)　上のリストに存在する機能表現の出現形の網羅的なリスト

現在利用可能な機能表現の見出し語のリストとして，グループ・ジャマシイのリスト (グループ・ジャマシイ 1998) や森田らのリスト (森田，松木 1989) などが存在する．しかし，それらを直接 (1) のリストとして利用することはできない．なぜならば，機能表現に関して，統一した見出し語選択方針が存在しないからである．例えば，上の2つは，異なる見出し語選択方針をとっている．前者においては，「にたいして」と「にたいする」は，両方とも見出し語である．その一方で，後者においては，「にたいして」のみが見出し語であり，「にたいする」は「にたいして」の派生として扱われている．機能表現辞書を編纂するにあたり，異なる複数の機能表現リストを併合するときに起こるこの問題を解決する必要がある．

自然言語処理システムは，実際の文章に現れる**出現形**を処理する必要があるので，上記 (2) のリストが必要となる．日本語母語話者は，機能表現の出現形からその見出し語を簡単に推測することができるので，人間のために編纂された機能表現の辞書には，出現形をすべて列挙しておく必要はない．実際，機能表現に関する，言語学や日本語教育学の文献は，このような記述形式をとっている (森田，松木 1989; グループ・ジャマシイ 1998)．計算機が利用することを想定した辞書を編纂する場合，機能表現の出現形を網羅する必要がある．

1章で挙げた問題，および上記で述べたことを考慮し，われわれは，編纂する辞書に次の3つの要件を設定した．

要件1　機能表現の出現形を網羅する見出し体系を持っていること

要件2　関連する機能表現間の関係が明示されていること

要件3　個々の機能表現に対して，文法情報や意味などが記述されていること

要件1を設定した理由は，すべての可能な機能表現の出現形を，計算機に誤りなく認識させたいからである．これが達成されれば，表記の問題と単位の問題を解決することができる．要件2を設定した理由は，この辞書を，異形式同意味の判定や言い換えに利用したいと考えているからである．これが達成されれば，外部情報源とのインタフェースの問題と異形式同意味の問題を解決することができる．要件3を設定した理由は，解析システムなどの自然言語処理システムに対して，個々の機能表現についての情報を提供することを想定しているからである．こ

れは，自然言語処理において利用されることを想定している辞書として，必須の条件である．

3.2 辞書の設計方針

前節の要件を満たす辞書を作成するにあたり，われわれは，設計方針を以下のように定めた．

見出し体系 9つの階層を持つ**階層構造**（次章参照）

この階層構造により，すべての機能表現の出現形を整理し，機能表現間の関係を明示する．

辞書の形式 XML形式

付加情報 以下に挙げる情報を記述する

左接続・右接続 隣に接続可能な形態素

意味カテゴリー 属する類義表現集合

「日本語表現文型」(森田，松木 1989) における意味分類を参考にして，同じような意味を持っている機能表現の類義表現集合として，89の意味カテゴリーを導入した．このうちのいずれかを記述する．

難易度 やさしい方からA1，A2，B，C，Fの5段階の難易度 (佐藤 2004)

「日本語能力試験出題基準」(国際交流基金，財団法人日本国際教育協会 2002) における「〈機能語〉の類」の級を参考にして，表現の分かりやすさに基づき，難易度を記述する．

文体 常体，敬体，口語体，堅い文体の4種類

核 表現の構成において中心的な核の形態素

例えば，「にたいして」に対して，「たいし」と記述する．

稀 使用が稀であることを示すマーク

例えば，「て呉れる」に対して，このマークを付与する．

例文 機能表現を含む文

否定表現 意味の観点から見た否定の表現

例えば，「なければならない」に対して，「なくてよい」と記述する．これを明記する理由は，機能表現の後ろに単純に「ない」を接続させた表現は，非文法的である場合があるからである．

慣用表現 機能表現を含むもの

例えば，「にたりない」に対して，「とるにたりない」と記述する．この情報は，慣用表現の一部である機能表現を，1単位であるとして誤検出してしまうことを防止することに利用することができる．

文献への参照 文献名および参照ページ

外部辞書の見出し語へのリンク 外部辞書における項目ID

見出し体系として採用した階層構造については，次章で詳しく説明する．辞書の形式としてXML形式を採用した理由は，次の2つである．

(1) 階層構造を表現するのに都合が良い
(2) 他の形式への変換が容易である

4 機能表現の階層構造

われわれは，2.3節で議論した，機能表現のさまざまな異形を扱うために，9つの階層を持つ階層構造を作成し，それを辞書の見出し体系として採用した．この階層構造は，3.1節で述べた要件1と2を満たす．具体的には，次のとおりである．

(1) 機能表現の出現形のリストとして，9つめの階層の機能表現集合を利用することができる
(2) それぞれの表現が持つIDを比較することにより，表現間の関係を知ることができる

4.1 9つの階層を持つ階層構造

われわれが作成した階層構造の一部を図2に示す．階層構造におけるL^0のルートノードは，すべての表現を統轄するダミーノードである．L^1の機能表現ノードは，辞書の見出し語に相当する．これは，最も抽象度の高い機能表現であると言える．一方，階層構造の葉に当たるL^9の機能表現ノードは，機能表現の出現形に相当する．これは，最も抽象度の低い機能表現であると言える．それらの間に存在する機能表現ノードは，中間の抽象度を持つ機能表現である．

階層構造の9つの階層について表3に示す．L^3からL^9が，2.3節で述べたそれぞれの異形のカテゴリーに対応する．

機能表現の階層構造を作成するにあたり，まず，われわれは，異形間の差異の大きさに基づいて，異形のカテゴリーに次の順番を定めた．

派生 (L^3) > 狭義の異形 (L^4, L^5, L^6) > 広義の活用 (L^7, L^8) > 表記のゆれ (L^9)

次に，単純な階層構造をつくるために，$L^4 \sim L^6$と$L^7 \sim L^8$を線形化した．狭義の異形 (L^4, L^5, L^6) において，機能語の交替 (L^4)，音韻的変化 (L^5)，とりたて詞の挿入 (L^6) という順番を定めた理由は，機能語の交替が，可能な音韻的変化に強く影響し，また，音韻的変化の有無が，とりたて詞の挿入に影響すると考えたからである．広義の活用 (L^7, L^8) において，活用 (L^7)，「です/ます」の有無 (L^8) という順番を定めた理由は，活用形の変化は，その表現が係ることのできる語の種類を変化させるが，「です/ます」の有無はそれとは独立であり，前者による異形は，後者による異形より，元の表現との差異が大きいと判断したからである．

そして，これらの階層の上に，次の2つの階層を定義した．

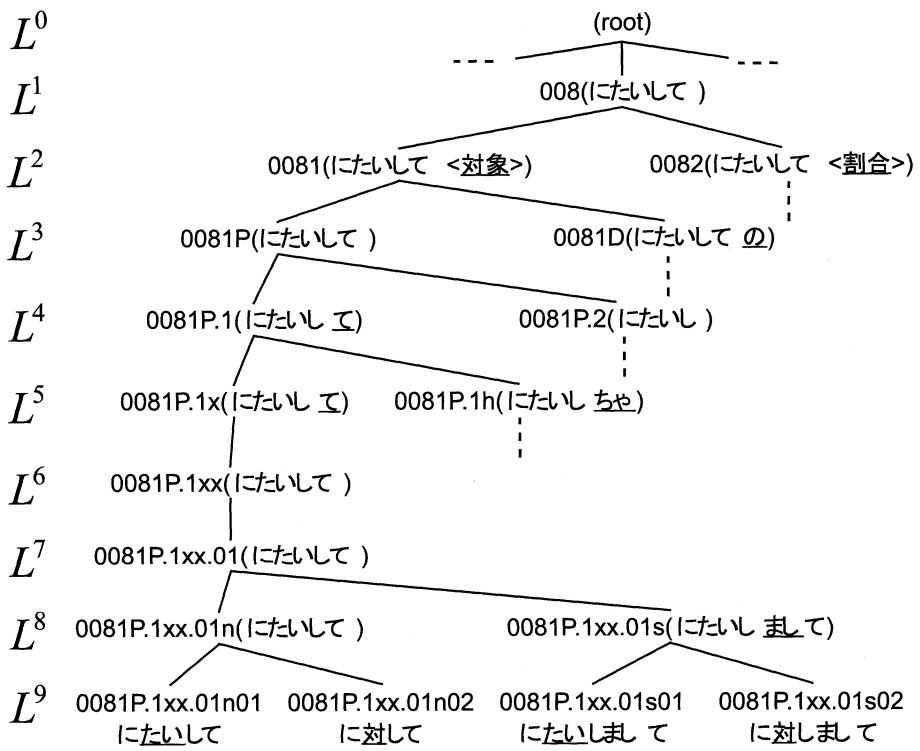

図 2　階層構造の一部

L^2　意味[1]

機能表現が，2つ以上の意味を持っている場合，この階層において，それらを区分する．例えば，「にたいして」は，2つの異なる意味を持っている．1つは，「彼は私にたいして優しい」において示されるような〈対象〉という意味であり，もう1つは，「一人にたいして5つ」において示されるような〈割合〉という意味である．この階層において，これらを区分する．

L^1　見出し語

辞書の見出し語に相当する．

この階層構造は，最も抽象度の高い形式の機能表現 (L^1) から最も抽象度の低い形式の機能表現 (L^9) までを含んでいるので，任意の形式の機能表現を，構造内に挿入することができる．逆に，この階層構造から，抽象度の異なる複数の機能表現リストを生成することができる．われわれの見出し語のリストは，L^1 の機能表現ノードの集合である．森田らのリスト (森田，松木

[1] 機能表現が持っている「意味・働き・役割」という概念に対して，言語学の文献では，主に，「意味」や「意味・用法」という用語が用いられている (永野 1953; 森田，松木 1989; 山崎，藤田 2001; 遠藤他 2003; 藤田，山崎 2006)．本論文では，この概念を表すのに「意味」を用いる．

表 3　9 つの階層

階層		ID		表現数
		文字種	長さ	
L^1	見出し語	数字	3	341
L^2	意味	数字	1	434
L^3	派生	英字（8 種類）	1	551
L^4	機能語の交替	数字	1	769
L^5	音韻的変化	英字（38 種類）	1	1,182
L^6	とりたて詞の挿入	英字（18 種類）	1	1,805
L^7	活用	数字	2	6,857
L^8	「です/ます」の有無	英字（2 種類）	1	9,705
L^9	表記のゆれ	数字	2	16,771

1989) におけるように，各々の見出し語は唯一の意味を持つという方針に従う場合，L^2 の機能表現ノードの集合を見出し語リストとして利用することができる．上記に加えて，各々の見出し語は唯一の機能を持つという方針に従う場合，L^3 の機能表現ノードの集合を見出し語リストとして利用することができる．一方，機能表現のすべての出現形のリストがほしいときには，L^9 の機能表現ノードの集合を利用することができる．

日本語機能表現辞書において，この階層構造を見出し体系として採用した．辞書の見出し体系に，すべての活用形を含めた理由は，辞書の各エントリーに活用型を記述する方法には，次の 2 つの問題があるからである．

(1) **活用体系に対して統一した見解が存在しない**

例えば，「なければならない」の末尾の「ない」を「ず」に置き換えた表現「なければならず」を，元の表現の活用形とみなす立場と，通常，助動詞「ない」の活用形に「ず」は含まれないため，全く異なる機能表現であるとみなす立場が存在する．

(2) **日本語として存在しない表現を生成してしまう可能性が高い**

複合辞は，動詞や助動詞と比べて，とることができる活用形が制限される傾向がある．例えば，「にほかならない」は，「ない」と同じだけの活用をするわけではなく，「*にほかならなから」，「*にほかならなかれ」，「*にほかならなけりゃ」といった表現は，日本語には存在しない[2]．辞書のエントリーに活用型を記述する方法を採用した場合，これらの非文法的な表現を，「にほかならない」の活用形として認めることになる．われわれは，解析だけではなく，生成や言い換えにおいてもこの辞書を利用することを想定しているので，これは大きな問題となる．

[2] 表現の先頭に付けた “*” は，その表現が非文法的であることを示す．

4.2 機能表現 ID

われわれは，機能表現の出現形（L^9 の機能表現）に対して，階層構造における位置を表す機能表現 ID を付与した．

機能表現 ID は，図 3 に示されるように，9 つの部分からなる．ID の各部分は，階層構造のそれぞれの階層における階層 ID である．階層 ID に用いる文字種とその長さを，表 3 の「ID」の欄に示す．L^3ID，L^5ID，L^6ID，L^8ID の一覧を，それぞれ，表 2，表 4，表 5，表 6 に示す．

機能表現 ID は，階層構造における位置を表しているので，ID を比較することにより，2 つの機能表現間の関係を容易に知ることができる．表 7 に，類似した ID を持つ 3 つの機能表現を示

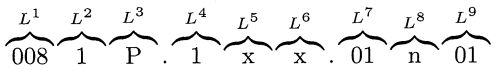

図 3 機能表現 ID の構成

表 4 音韻的変化と L^5ID （38 種類）

L^5ID	縮約	脱落	促音化・撥音化	有声音化
a/A	てあ			（なし）/頭音
b	れば			
c/C	てし			（なし）/頭音
d				た
e				て
g/G		が		（なし）/頭音
h/H	は			（なし）/頭音
i/I		い（っ）		（なし）/頭音
j/J	てしま			（なし）/頭音
k	しか			
l/L		い（っ） and ろ		（なし）/頭音
m/M	も			（なし）/頭音
n/N			撥音化	（なし）/頭音
o/O	てお			（なし）/頭音
p/P	は		撥音化	（なし）/頭音
q/Q			促音化	（なし）/頭音
r/R		ろ		（なし）/頭音
s/S	しよ			（なし）/頭音
v				く
w			促音化 and 撥音化	
x				
y	ければ			
z/Z	しよ and も			（なし）/頭音

表 5　とりたて詞と L^6ID （18 種類）

L^6ID	とりたて詞	L^6ID	とりたて詞
h	は	i	のみ
m	も	c	しか
e	さえ	o	こそ
d	でも	n	など
s	すら	g	なんか
q	だって	t	なんて
a	まで	r	くらい
l	だけ	k	か
b	ばかり	x	（挿入なし）

表 6　「です/ます」の有無と L^8ID （2 種類）

L^8ID	「です/ます」の有無
n	無
s	有

表 7　類似した ID を持つ機能表現

	ID	機能表現
(1)	0081P.1xx.01n01	にたいして
(2)	0081P.1hx.01n01	にたいしちゃ
(3)	0091P.1xx.01n01	について

す．(1) の ID と (2) の ID の差異は，8 文字めの “x” と “h” である．これらは L^5ID であるので，「にたいして」と「にたいしちゃ」は，同じ見出し語に対する音韻的異形の関係にあることが分かる．それに対して，(1) の ID と (3) の ID は，最初の 3 文字が異なっている．これらは L^1ID であるので，「にたいして」と「について」は，全く異なる機能表現であるということが分かる．

なお，図 2 に示されるように，L^1 から L^8 の機能表現に対しても，階層構造における位置を表す ID を付与した．

5　機能表現辞書の編纂

5.1　編纂手順

われわれは，前章で説明した階層構造を見出し体系として持つ日本語機能表現辞書を編纂した．

機能表現辞書の編纂は，図 2 に示されるような階層構造の木を徐々に成長させる過程に相当する．その理由は，機能表現の完全な見出し語のリストも，すべての可能な出現形のリストも利用可能ではないからである．言語学や日本語教育学の文献に収録されている機能表現を，1 つずつ，階層構造に挿入する過程を繰り返し，辞書を徐々に大きくしていくことになる．

文献から得た 1 つの機能表現に対する編纂作業は，次のとおりである．

(1)　その表現に対するノードを作り，それを階層構造の適切な位置に挿入する
(2)　表現の形態に応じたテンプレートを用いて，その機能表現の異形と思われる表現を含む部分木を生成し，それを階層構造の適切な位置に挿入する
(3)　過生成されてしまった，実際に存在しない表現を消去する
(4)　テンプレートでは生成できなかった特別な異形を追加する
(5)　3.2 節に挙げた付加情報を記述する

言語学や日本語教育学の文献における大部分の見出し語は，階層構造における L^2 の機能表現

ノードに対応する．残りの見出し語は，L^4 や L^5 の機能表現ノードに対応する．見出し語を階層構造に挿入するときには，その見出し語に対するノードだけではなく，必要に応じて，上位のノードも作成する．

上記の (2) で用いるテンプレートは，全部で 10 種類ある．例えば，「に」+動詞のテ形+「て」という形態の機能表現に対して適用することができるテンプレートがある．例えば，このテンプレートを「にかんして」に適用した場合，L^2 の機能表現ノードをルートとして持つ，階層構造の部分木が生成される．この部分木は，L^3，L^4，L^5，L^8 に，「にかんする」，「にかんし」，「にかんしちゃ」，「にかんしまして」など，「にかんして」の異形に対応する機能表現ノードを持ち，葉の部分 (L^9) に，「にかんして」の出現形に対応する機能表現ノードを持つ．

5.2 機能表現辞書の現状

2.2 節で述べたように，本研究では，機能表現であるかどうかの判断が難しい表現は扱わず，定評のある文献において機能表現であると認められているもののみを扱う．われわれは，次の 2 つのリストに含まれるすべての機能表現を，機能表現辞書に収録した．

(1) **「日本語表現文型 用例中心・複合辞の意味と用法」** (森田，松木 1989) に収録されている，助詞と同様の働きをする表現（ただし，並立助詞の働きをするものは除く）と助動詞と同様の働きをする表現，計 412 表現

(2) **「使い方の分かる類語例解辞典 新装版」** (遠藤他 2003) の助詞・助動詞解説編に収録されている助詞，助動詞およびその連接形，計 368 表現

これらの文献を選んだ理由は，次のとおりである．

- 「日本語表現文型」は，数百の複合辞の意味と用法について詳細に解説した最初の文献であり，複合辞の用例を数多く収録している．この文献のリストは，基本的な複合辞をすべて網羅しているため，機能表現辞書に最初に収録する機能表現集合として最適であると考えたからである．
- 「使い方の分かる類語例解辞典」は，その助詞・助動詞解説編において，78 のカテゴリーを持つ，機能表現のシソーラスを収録している．このシソーラスは，機能語と複合辞の両方を含んでいるため，機能表現辞書に機能語を収録するにあたり，非常に有用であると考えたからである．

階層構造の各階層における機能表現の数を，表 3 の「表現数」の欄に示す．見出し語に相当する L^1 の機能表現の数は 341 であり，出現形に相当する L^9 の機能表現の数は 16,771 である．

3.2 節で述べた付加情報は，階層構造の葉にあたる L^9 の機能表現ノードにではなく，適切な中間ノードに記述した．例えば，意味カテゴリーや難易度は，L^2 の機能表現ノードに，文体は，L^5 の機能表現ノードに記述した．L^9 の機能表現ノードを含む下位のノードは，階層構造の特徴を利用して，それらの情報を継承するが，必要に応じて，そこに異なる情報を記述することに

より，継承された情報を上書きすることができる．このような記述法をとることにより，付加情報と形式の間の関係を明確にすることができる．例えば，文体は，L^5 と L^7 の機能表現ノードにのみ記述しているので，表記 (L^9) とは無関係であることを示すことができる．機能表現辞書における意味カテゴリー，難易度，文体の分布を，それぞれ，表 8，表 9，表 10 に示す．また，機能表現を構成する語の数の分布を表 11 に示す．ここで，構成語数が 1 のものには，機能語以外に，「ため」，「折」など，形式的には名詞一語である表現や，「たげる」，「ちゃう」など，縮約された表現が含まれている．

外部辞書の見出し語へのリンクとして，L^2 もしくは L^5 の機能表現ノードに，次の 2 つのリストにおける項目の ID を記述した．

(1) **「現代語複合辞用例集」**(山崎，藤田 2001)
日本語において代表的な複合辞に対して，その用例と解説を記載している．収録されている複合辞は，「日本語表現文型」のほぼ部分集合であり，収録表現数は 129 である．それぞれの表現に対して，表 12 に示される 3 文字の項目 ID が付与されている．

(2) **「使い方の分かる類語例解辞典 新装版」**(遠藤他 2003)
この節の冒頭で説明した辞典である．収録表現数は 368 である．
シソーラスの各々のカテゴリーには，見出し語の集合と関連語の集合が存在する．各カ

表 8　意味カテゴリーの分布

L^2 の表現数	意味カテゴリー　（意味カテゴリー数）
15	逆接確定 (1)
14	推量，強調，否定，状況 (4)
13	理由 (1)
12	— (0)
11	同時性，対象 (2)
10	感嘆，限定，自然発生 (3)
9	依頼，疑問，並立，話題 (4)
8	意志，願望，逆接仮定，順接仮定，想外 (5)
7	当為，勧め，継起，仲介 (4)
6	伝聞，起点，順接確定，添加 (4)
5	判断，不可能，不必要，対比，立場，極端例，主体 (7)
4	許可，不許可，不可避，勧誘，継続，事後，終点，相関，付帯，放置 (10)
3	可能，順接限定，非限定，因状況，回想，完了，基準，根拠，目的，同格，不満，比況，も観点，内-授与，着継続，程度，相手 (17)
2	範囲，否定意志，定義，割合，相応，事前，順接必要，発継続，不均衡，習慣，状態，内-受益，他-授与，目標，は観点，不明確，無意味 (17)
1	否定推量，回避，傾向，経験，最中，場合，適当，反復，無視，名詞化 (10)

表 9 難易度の分布

階層	A1	A2	B	C	F	計
L^2	81	86	143	92	32	434
	(19%)	(20%)	(33%)	(21%)	(7%)	(100%)
L^3	99	107	191	117	37	551
	(18%)	(19%)	(35%)	(21%)	(7%)	(100%)
L^4	106	160	305	158	40	769
	(14%)	(21%)	(40%)	(20%)	(5%)	(100%)
L^5	141	303	430	251	57	1,182
	(12%)	(26%)	(36%)	(21%)	(5%)	(100%)
L^6	179	588	569	363	106	1,805
	(10%)	(33%)	(31%)	(20%)	(6%)	(100%)
L^7	441	2,857	1,812	1,309	438	6,857
	(6%)	(42%)	(27%)	(19%)	(6%)	(100%)
L^8	538	4,187	2,675	1,834	471	9,705
	(5%)	(43%)	(28%)	(19%)	(5%)	(100%)
L^9	684	6,964	5,129	3,329	665	16,771
	(4%)	(41%)	(31%)	(20%)	(4%)	(100%)

表 10 文体の分布

階層	常体	敬体	口語体	堅い文体	計
L^5	782	42	311	47	1,182
	(66%)	(4%)	(26%)	(4%)	(100%)
L^6	1,209	166	338	92	1,805
	(67%)	(9%)	(19%)	(5%)	(100%)
L^7	3,458	789	1,918	692	6,857
	(50%)	(12%)	(28%)	(10%)	(100%)
L^8	3,340	2,815	2,176	1,374	9,705
	(35%)	(29%)	(22%)	(14%)	(100%)
L^9	5,531	5,379	3,904	1,957	16,771
	(33%)	(32%)	(23%)	(12%)	(100%)

テゴリーには，2桁の数字からなるIDが設定されているが，カテゴリー内の各語に対しては，IDが設定されていない．そこで，われわれは，カテゴリーごとに，そこに存在する語に対して，表13のように下位IDを設定し，カテゴリーのIDと合わせて，“29N01”や“75R03”のような5文字の項目IDを定めた．

5.3 異形の被覆率の評価

機能表現辞書を評価する観点として，次の2点が考えられる．

表 11　構成語数の分布

階層	1	2	3	4	5	6	7	8	計
L^1	64	101	120	40	14	2	—	—	341
	(19%)	(29%)	(35%)	(12%)	(4%)	(1%)			(100%)
L^2	105	126	138	49	14	2	—	—	434
	(24%)	(29%)	(32%)	(11%)	(3%)	(1%)			(100%)
L^3	119	146	177	90	16	3	—	—	551
	(22%)	(26%)	(32%)	(16%)	(3%)	(1%)			(100%)
L^4	131	215	247	136	31	9	—	—	769
	(17%)	(28%)	(32%)	(18%)	(4%)	(1%)			(100%)
L^5	180	342	385	204	49	22	—	—	1,182
	(15%)	(29%)	(33%)	(17%)	(4%)	(2%)			(100%)
L^6	180	342	673	411	138	34	27	—	1,805
	(10%)	(19%)	(37%)	(23%)	(8%)	(2%)	(1%)		(100%)
L^7	289	774	2,851	2,004	624	174	117	24	6,857
	(4%)	(11%)	(42%)	(29%)	(9%)	(3%)	(2%)	(0%)	(100%)
L^8	292	837	3,132	3,342	1,377	420	256	49	9,705
	(3%)	(9%)	(32%)	(34%)	(14%)	(4%)	(3%)	(1%)	(100%)
L^9	347	1,299	5,597	5,729	2,285	820	580	114	16,771
	(2%)	(8%)	(33%)	(34%)	(14%)	(5%)	(3%)	(1%)	(100%)

表 12　「現代語複合辞用例集」(山崎，藤田 2001) の項目 ID

大分類	項目 ID
助詞的複合辞	A01，A02，…，A83
助動詞的複合辞	B01，B02，…，B42
参考表現	R01，R02，R03，R04

表 13　「使い方の分かる類語例解辞典 新装版」(遠藤他 2003) の語の下位 ID

語	下位 ID
見出し語	N01，N02，N03，…
関連語	R01，R02，R03，…

(1)　機能表現をどのくらい収録しているか

(2)　応用システムにおいてどれほど有用であるか

ここでは，前者の評価を行なう．一般に，辞書に収録されている機能表現の被覆率を評価することは難しい．なぜならば，2.2 節で述べたように，機能表現と呼べる表現の範囲や機能表現の単位に対して，統一した見解が存在しないからである．それゆえ，機能表現集合の母集団が不明であり，単純に被覆率を計算することができない．被覆率の評価の際には，これらに起因す

る問題を適切に扱う必要がある．

収録している機能表現の被覆率を評価する場合，次の 2 点を評価する必要がある．

(1) 見出し語の被覆率

(2) 異形（の出現形）の被覆率

2.2 節で述べた方針に基づき，われわれは，前者は，辞書構築時に利用した，定評のある文献によって保証されていると考える．よって，ここでは，後者の評価のみを行なう．本研究では，既存の機能表現リストと比較することにより，機能表現辞書に収録されている異形の被覆率を評価した．

評価には，比較対象として，自然言語処理の分野において唯一われわれが利用可能であり，かつ，大規模な機能表現リストを収録している首藤の文献 (首藤 1980) を用いた[3]．首藤の文献は，独自の文節モデルと日本語の解析システムについて述べたものであり，その付録に，付属語的表現，接続詞的表現，接尾語的表現，副詞的表現，連体詞的表現の一覧を記載している．付属語的表現とは，文節において構造的意味情報を担う表現のことであり，おおきく，文節間の関係を指示する表現（関係表現）と話し手の判断や叙述の仕方などを表す表現（助述表現）に分類される．前者は，ほぼ，われわれの助詞型機能表現に相当し，後者は，ほぼ，われわれの助動詞型機能表現に相当する．われわれが機能表現辞書編纂時に利用した 2 つの文献 (森田，松木 1989; 遠藤他 2003) は，接続詞型機能表現をほとんど扱っていなかったので，われわれの機能表現辞書は，接続詞型機能表現をほとんど収録していない．このような理由により，異形の被覆率の評価という観点から，接続詞型以外の機能型の表現のみを比較対象とし，機能表現辞書の機能表現と首藤の付属語的表現を比較することにした．

首藤の付属語的表現は，表 14 のような形式で記載されている．左の欄の R_{NP1} は，基本的な文法カテゴリーであり，この場合は，「名詞文節から述語文節へと係る格助詞的表現」を表す．右の欄の R19 は，連接規則を精密化するために導入された文法カテゴリーの下位分類である．真ん中の欄に，これらのカテゴリーに属する付属語的表現が記述される．それが関係表現の場合，表 14 のように，表現の後に意味を表す記号（例えば，〈対象 1〉＋〈領域的〉）が続く．こ

表 14 首藤の付属語的表現の記載形式

R_{NP1}	···，にさいして（〈時点〉+/），にたいして（〈使役文の動作者〉+/，〈対象 4〉+/，〈対象 6〉+/，〈抽象的場〉+/，〈対立〉+/），にわたって（〈対象 1〉＋〈領域的〉，〈抽象的場〉＋〈領域的〉，〈受動者〉＋〈領域的〉，〈依頼者〉＋〈領域的〉，〈場所〉＋〈領域的〉），···	R19

[3] 3.1 節で述べたように，言語学や日本語教育学の文献は，人間が読むことを想定して書かれたものであるので，異形を網羅することを目指しておらず，異形をあまり収録していない．それゆえ，これらの分野の文献は，比較対象として適切ではない．

のリストは，次のような5つの特徴を持つ．

(1) “ASU”，“ASERU”，“ERU”，“ARERU” の4表現を除き，表現は，すべてひらがな表記であり，漢字表記のものは存在しない

(2) 「でもよい」や「こともできる」など，音韻的変化やとりたて詞の挿入による異形が含まれるが，「です/ます」を含む異形は存在しない

(3) 活用による異形は，「なければならず」や「かもしれず」など，表現の末尾の「ない」を「ず」に置き換えたもののみである

(4) 「たにちがいない」や「ないかもしれなかった」など，われわれが複数の機能表現からなるとみなす表現が存在する

(5) 「じゅんに」や「けいかしてから」など，われわれが機能表現とみなさない表現が含まれる

計算機処理を可能にするために，文法カテゴリーと意味記号を無視し，このリストに含まれる付属語的表現を人手で入力し，電子データを作成した．このとき，アルファベット表記である上記の4表現は除いた．入力したデータに含まれる表現数は，937であった．以下，このデータを首藤リストと呼ぶ．

機能表現辞書に含まれる異形の被覆率を評価する場合，異形がすべて展開された L^9 の機能表現集合を用いるのが適切である．一方，首藤リストの特徴から，次の4つのことが言える．

(1) 首藤リストには，漢字表記の異形が存在しないため，L^9 の機能表現集合に対する比較結果と，L^8 の機能表現集合に対する比較結果は同じになる

(2) 首藤リストには，「です/ます」の有無による異形が存在しないため，L^8 の機能表現集合に対する比較結果と，L^7 の機能表現集合に対する比較結果は同じになる

(3) 首藤リストには，活用による異形は，表現の末尾の「ない」を「ず」に置き換えた表現のみであるので，L^7 の機能表現集合に対する比較結果と，L^6 の機能表現集合に対する比較結果の差分は，それに関する少数のもののみであることが予想される

(4) 首藤リストには，とりたて詞の挿入による異形が含まれているため，それを扱っていない L^1 から L^5 の機能表現集合を用いることは適切ではない

これらの理由により，本評価においては，L^6，L^7 の機能表現集合と首藤リストを比較した．比較にあたっては，首藤リストが持つ特徴を考慮し，首藤リストの表現を次の5種類に分類した．

(A) L^6 の機能表現集合に含まれる

(B) L^6 の機能表現集合には含まれないが，L^7 の機能表現集合に含まれる

(C) 複数の L^7 の機能表現から構成されている

(D) 機能表現辞書に含まれていない

(E) 機能表現ではない

分類手順は，次のとおりである．

(1) 首藤リストから，L^6 の機能表現と文字列が完全に一致するものをすべて抽出する（上記の (A) に相当）

(2) 残ったリストから，L^7 の機能表現と文字列が完全に一致するものをすべて抽出する（上記の (B) に相当）

(3) 残ったリストの表現を，次の 3 つのいずれかに人手で分類する

(i) 複数の L^7 の機能表現から構成されている（上記の (C) に相当）

(ii) 機能表現ではない（上記の (E) に相当）

(iii) それ以外（上記の (D) に相当）

分類結果を表 15 に示す．首藤リストの表現のうち，(A)，(B)，(C) の表現は，機能表現辞書が被覆していると言える．一方，(E) の表現は，われわれは，機能表現とはみなさないので，比較対象から除外すべきである．よって，首藤リストに対する，機能表現辞書の被覆率を，次の式で算出する．

$$\text{被覆率 (\%)} = \frac{\text{(A)} + \text{(B)} + \text{(C)}}{\text{(A)} + \text{(B)} + \text{(C)} + \text{(D)}} \times 100$$

表 15 の値を用いて計算すると，これは，84% (530/630) であった．残りの 16%を占める (D) の表現はすべて，「にもとづいて」や「とひかくして」など，機能表現であるかどうかの判断が難しいものであった[4]．3 表現を除いて，これらの表現は，もし階層構造へ挿入するならば，その挿入に，新しい L^1 の機能表現ノードを作る必要がある表現であった．それゆえ，機能表現辞書は，既存の機能表現リストに含まれる異形をほとんどすべて収録していると言うことができる．

表 15 首藤リストの表現の分類結果

		表現数
(A)	L^6 の機能表現集合に含まれる	286
(B)	L^6 の機能表現集合には含まれないが，L^7 の機能表現集合に含まれる	14
(C)	複数の L^7 の機能表現から構成されている	230
(D)	機能表現辞書に含まれていない	100
(E)	機能表現ではない	307
計		937

[4] これらの多くは，定型的な英訳（例えば，「にもとづいて」は “based on”，「とひかくして」は “compared to”）を持つ．後続処理に機械翻訳を想定した場合，解析システムの辞書にこれらの表現を登録すると都合がよい．比較に利用した首藤リストは，このような考え方に基づいて作成された可能性が高いと思われる．

6　関連研究

言語学，日本語教育学，自然言語処理の3つの分野における，機能表現（特に，複合辞）の扱いについて述べる．

6.1　言語学（日本語学）

複合辞というとらえ方を初めて提唱したのは，国立国語研究所の資料 (山崎，藤田 2001) によると，永野 (永野 1953) である．永野は，語源的・構造的にはさらにいくつかの語に分解できるが，単なる部分の合成以上の「一まとまりの意味を持っているものと見てよい」連語形式の助詞相当表現の存在を指摘し，これを複合助詞と呼んだ．そして，同様の基準で複合助動詞，複合感動詞，複合接続詞についても考え，複合助詞とこれらを合わせて，複合辞と呼んだ．

森田ら (森田，松木 1989) は，複合辞に関する大量の用例を収集し，数百の複合辞の意味と用法について詳細に分析している．また，この研究を受け，国立国語研究所は，代表的な複合辞を選定し，それらの用例集を作成した (山崎，藤田 2001)．

言語学においては，現在も，複合辞の研究が活発に続けられている (藤田，山崎 2006)．

6.2　日本語教育学

日本語教育学において，複合辞は，文法項目として重要視されている．例えば，日本語能力試験1，2級の文法問題を解くためには，さまざまな種類の複合辞について正しく理解している必要がある (国際交流基金，財団法人日本国際教育協会 2002)．そして，この理解を助けるために，日本語学習者のための，日本語文法の辞典は，複合辞を見出しに立て，それらについて詳しく解説していることが多い (Makino and Tsutsui. 1986, 1995; グループ・ジャマシイ 1998)．

6.3　自然言語処理

自然言語処理において，複合辞は，それらを一まとまりの意味の塊として扱う必要があることから，特に，機械翻訳において重要視されてきた．首藤ら (首藤 1980; Shudo et al. 1980) は，機械翻訳への入力とするために，概念を表す内容語と機能的な付属語列からなる拡張文節という考え方を導入し，そこで機能語に相当する1単位として複合辞を扱っている．現在，彼らは，日本語において，2500の機能表現を収集し，それらを意味に基づいて分類している (Shudo et al. 2004)．しかしながら，異形についての大規模な整理は行なわれておらず，彼らの辞書は，五十音順以外に特別な構造を持っていないようである．

EDR 日本語単語辞書 (日本電子化辞書研究所 2001) には，助詞相当語82表現，助動詞相当語49表現が登録されているが，異形に関する情報は記載されていない．

兵藤ら (兵藤他 2000) は，2つの層を持つ日本語機能表現の辞書を提案している．第一の層に

は，375の項目があり，これらの項目から，第二の層において，自動的に13,882の可能な出現形が生成される．この辞書は，表現のある部分に対して交換可能な文字列を列挙しているだけであり，2つの異なる表現間の関係についての情報を何も提供しない．

これらの辞書が機能表現の異形を適切に扱っていないのに対して，われわれの辞書は，階層構造に基づいて機能表現の異形を整理しており，2つの表現間の関係について，「音韻的異形」や「表記のゆれ」といった情報を提供することができる．

日本語話し言葉コーパスにおいては，助詞相当句29表現と助動詞相当句37表現が，長単位の見出し語として扱われている(小椋他 2004)．それらの表現は，丁寧形や異形態などの観点から，前者は80，後者は92の表現に細分されている．

土屋ら(土屋他 2006)は，複合辞の用例データベースを作成するにあたり，「現代語複合辞用例集」(山崎，藤田 2001)に記載されている複合辞123項目（見出し語に相当）に対して異形を展開し，細分した337小項目の表現を用例収集の対象としている．彼らは，助詞の交替や文体などの観点から異形を分類し，それぞれの小項目に対して8文字のIDを付与している．

これらの研究の機能表現リストは，小規模なものである．一方，われわれの辞書は，見出し語で341，出現形で16,771の機能表現を分類整理している．

7 おわりに

本論文では，われわれが作成した，自然言語処理のための日本語機能表現辞書について報告した．この辞書は，機能表現のさまざまな異形を扱うために，見出し体系として，9つの階層からなる階層構造を持つ．現在，この辞書には，341の見出し語と16,771の出現形が収録されている．既存の機能表現リストと比較した結果，各々の見出し語に対して，ほぼすべての異形が網羅されていることが分かった．

われわれの機能表現辞書は，日本語文の解析，生成，言い換えなど，さまざまな自然言語処理タスクにおいて利用することができる．例えば，この辞書はほとんどすべての異形を収録しているので，機能表現解析システムの検出被覆率の向上に役立つ．また，辞書中のすべての機能表現が意味カテゴリーの情報を持つので，この辞書を利用することにより，機能表現を類義表現に言い換えるシステムを容易に構築することができると思われる．機能表現辞書を用いた応用システムの性能からこの辞書を評価することは，今後の課題である．

参考文献

浅原正幸，高橋由梨加，松本裕治 (2005). “異表記同語情報を付与した辞書の整備.” 言語処理学会 第11回年次大会発表論文集, pp. 604–607.

遠藤織枝，小林賢次，三井昭子，村木新次郎，吉沢靖（編） (2003). 使い方の分かる類語例解辞典 新装版. 小学館.

藤田保幸，山崎誠（編） (2006). 複合辞研究の現在. 和泉書院.

グループ・ジャマシイ（編） (1998). 教師と学習者のための日本語文型辞典. くろしお出版.

橋本力，佐藤理史，宇津呂武仁 (2006a). “自動検出のための慣用句の分類と語彙的情報.” 言語処理学会 第12回年次大会発表論文集, pp. 825–828.

橋本力，佐藤理史，宇津呂武仁 (2006b). “依存構造照合に基づく慣用句自動検出.” 言語処理学会 第12回年次大会発表論文集, pp. 829–832.

兵藤安昭，村上裕，池田尚志 (2000). “文節解析のための長単位機能語辞書.” 言語処理学会 第6回年次大会発表論文集, pp. 407–410.

国際交流基金，財団法人日本国際教育協会（編） (2002). 日本語能力試験出題基準【改訂版】. 凡人社.

黒橋禎夫，河原大輔 (2005). “日本語形態素解析システム JUMAN version 5.1.” http://nlp.kuee.kyoto-u.ac.jp/nl-resource/juman.html.

Makino, S. and Tsutsui., M. (1986). *A Dictionary of Basic Japanese Grammar.* The Japan Times.

Makino, S. and Tsutsui., M. (1995). *A Dictionary of Intermediate Japanese Grammar.* The Japan Times.

森田良行，松木正恵 (1989). 日本語表現文型 用例中心・複合辞の意味と用法. アルク.

永野賢 (1953). “表現文法の問題-複合辞の認定について-.” 金田一博士古稀記念論文集刊行会（編）, 金田一博士古稀記念言語民族論叢. 三省堂. 「永野 賢 (1970). 伝達論にもとづく日本語文法の研究. 東京堂出版」に再録.

日本電子化辞書研究所 (2001). “EDR 電子化辞書 2.0 版 仕様説明書 第2章 日本語単語辞書.”.

沼田善子 (1986). “とりたて詞.” 奥津敬一郎，沼田善子，杉本武（編）, いわゆる日本語助詞の研究, 2章. 凡人社.

小椋秀樹，山口昌也，西川賢哉，石塚京子，木村睦子 (2004). 『日本語話し言葉コーパス』の形態論情報の概要 ver. 1.0. 国立国語研究所.

尾嶋憲治，佐藤理史，宇津呂武仁 (2006). “日本語慣用句用例データベースの構築法.” 言語処理学会 第12回年次大会発表論文集, pp. 456–459.

佐藤理史 (2004). “異表記同語認定のための辞書編纂.” 情報処理学会研究報告 2004-NL-161, pp. 97–104.

首藤公昭 (1980). 文節構造モデルによる日本語の機械処理に関する研究. 福岡大学研究所報第45号.

Shudo, K., Narahara, T., and Yoshida, S. (1980). “Morphological Aspect of Japanese Language

Processing." In *Proceedings of the 8th COLING*, pp. 1–8.

Shudo, K., Tanabe, T., Takahashi, M., and Yoshimura, K. (2004). "MWEs as Non-propositional Content Indicators." In *Proceedings of the 2nd ACL Workshop on Multiword Expressions: Integrating Processing (MWE-2004)*, pp. 32–39.

土屋雅稔，宇津呂武仁，松吉俊，佐藤理史，中川聖一 (2006). "日本語複合辞用例データベースの作成と分析." 情報処理学会論文誌, **47** (6), pp. 1728–1741.

山崎誠，藤田保幸（編） (2001). 現代語複合辞用例集. 国立国語研究所.

略歴

松吉　俊：2003 年京都大学理学部卒業．2005 年同大学院情報学研究科修士課程修了．現在，同大学院情報学研究科博士後期課程在学中．自然言語処理の研究に従事．

佐藤　理史：1983 年京都大学工学部電気工学第二学科卒業．1988 年同大学院工学研究科博士後期課程電気工学第二専攻研究指導認定退学．京都大学工学部助手，北陸先端科学技術大学院大学情報科学研究科助教授，京都大学大学院情報学研究科助教授を経て，2005 年より名古屋大学大学院工学研究科電子情報システム専攻教授．工学博士．自然言語処理，情報の自動編集等の研究に従事．

宇津呂武仁：1989 年京都大学工学部電気工学第二学科卒業．1994 年同大学大学院工学研究科博士課程電気工学第二専攻修了．京都大学博士（工学）．奈良先端科学技術大学院大学情報科学研究科助手，豊橋技術科学大学工学部情報工学系講師，京都大学情報学研究科知能情報学専攻講師を経て，2006 年より筑波大学大学院システム情報工学研究科知能機能システム専攻准教授．自然言語処理の研究に従事．

（2007 年 3 月 9 日　受付）
（2007 年 5 月 18 日　再受付）
（2007 年 6 月 23 日　採録）